Language Learning 56:3, September 2006, pp. 463–506

Effects of Stimulus Characteristics and Background Music on Foreign Language Vocabulary Learning and Forgetting

Annette M. B. de Groot
University of Amsterdam

This study examined the effects of three stimulus variables and background music on paired-associate learning of foreign language (FL) vocabulary. The stimulus variables were the frequency and concreteness of the native language (L1) words and the (phonotactical) typicality of the FL words. Sixty-four L1-FL pairs were presented for learning six times, followed by a recall test after the second, fourth, and sixth learning round. A fourth recall test took place 1 week later. Typical FL words, FL words paired with frequent L1 words, and FL words paired with concrete L1 words were learned better than atypical FL words and FL words paired with infrequent and abstract L1 words, respectively. More learning occurred in the music condition than in the silent condition. The results are interpreted in terms of differences between memory representations of L1 words, differences in the phonological coding enabled by the FL words, and individual learner differences.

I am greatly indebted to Rosanne van den Brink, who collected the data of this study in partial fulfilment of her master's thesis (Van den Brink, 2000). Furthermore, I thank Marina Pool for drawing my attention to a number of studies on the effect of background music on human performance in cognitive tasks, and Don van Ravenzwaaij for assistance in analyzing the data.

Correspondence concerning this article should be addressed to Annette M. B. de Groot, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB, Amsterdam, The Netherlands. Internet: a.m.b.degroot@uva.nl

Native speakers can better understand ungrammatical utterances with accurate vocabulary spoken by foreign language (FL) learners of their language than utterances with accurate grammar but inaccurate vocabulary (Boyd Zimmerman, 1997). Obviously then, vocabulary knowledge is of crucial importance to the FL learner. Yet, vocabulary has until recently received relatively little attention by both FL teachers and FL researchers (Boyd Zimmerman, 1997; but see Bogaards & Laufer, 2004, a volume that illustrates the growing interest in the study of FL vocabulary teaching/learning).

A reason why vocabulary learning has been ignored until recently might have been the apparent impossibility of the task due to the vastness of the lexicon: A language typically contains tens of thousands of words, most of which have multiple meanings, and for each of which eight types of information have to be learned: phonological, orthographic, syntactic, morphological, pragmatic, articulatory, idiomatic, and semantic (Schreuder, 1987). However, the task of learning an adequate FL vocabulary can be reduced to manageable proportions if the curriculum designers take into account a well-known aspect of the statistics of any language, namely that words differ greatly from one another in their frequency of occurrence in natural language use (and, as demonstrated by Zipf, 1968, in an orderly way). If FL vocabulary teaching focuses on the words that occur most frequently, a reasonably large coverage can be attained with relatively small numbers of known words. Laufer (1992) observed that an adequate level of comprehension in the FL was reached by learners whose vocabulary size covered 95% of the words in a text, a state of affairs that Nation (1993) argued exists if the learner has a vocabulary size of the 5,000 most frequent lexical items, equaling about 3,000 word families. The important implication is that, after all, including direct vocabulary training of carefully selected words in an FL curriculum is a feasible option. The speed with which an adequate vocabulary is attained might be increased further if the curriculum would, in addition to usage frequency, take into account other factors known to affect the learning process.

Among the factors that affect FL vocabulary learning are specific characteristics of the L1 words for which an FL translation is to be learned (e.g., de Groot & Keijzer, 2000) or of the FL words themselves (e.g., Laufer, 1997), the relation (similarity, difference, or a mixture of both similarity and difference) between the first or native language (L1) lexical forms and the corresponding FL lexical forms (e.g., Granger, 1993; Hall, 2002; Laufer & Eliasson, 1993; Lotto & de Groot, 1998; Talamas, Kroll, & Dufour, 1999), the overall similarity of the L1 and the FL to be learned (e.g., Ard & Homburg, 1992), the teaching method that is employed (e.g., Hogben & Lawson, 1997; Lawson & Hogben, 1998; Lotto & de Groot; Van Hell & Candia Mahn, 1997), the specifics of the learning environment (e.g., Felix, 1993), and individual learner differences such as the seven categories of learner differences identified by Ellis (1994): age, aptitude, motivation, learning style, beliefs, affective states, and personality. In addition, learning success might depend on interactions between a number of these factors—for instance, on a match between learner strategy and a given learning environment (Segalowitz, 1997) or between a particular teaching method and whether the learners are experienced FL learners (e.g., Van Hell & Candia Mahn).

Earlier Related Studies

This article reports on the latest of a series of studies performed in our laboratory that all tried to contribute to our understanding of what factors influence FL vocabulary learning. Importantly, all of these studies only dealt with a very initial stage of FL vocabulary learning, where (a representation of) the name of the FL word to be learned is attached onto a representation of the corresponding L1 word that already existed in memory at the onset of learning (see, e.g., the "parasitic" model of vocabulary development of Hall, 2002, and Hall & Ecke, 2003, and the developmental model proposed by Kroll & Stewart, 1994). These earlier studies investigated the effects of stimulus characteristics (de Groot & Keijzer, 2000; Lotto & de Groot, 1998; Van Hell &

Candia Mahn, 1997), learning method (Lotto & de Groot; Van Hell & Candia Mahn), and the amount of previous experience in FL vocabulary learning (Van Hell & Candia Mahn). Two of these studies (de Groot & Keijzer; Lotto & de Groot) employed "paired-associate" learning, which is a simple, yet very effective, method of direct (initial) vocabulary teaching (or, from the learner's viewpoint, of vocabulary *learning*): During learning, pairs of stimuli are presented; one of them is a word in the target FL (natural or artificial) and the second is the corresponding L1 word ("word-association" learning) or a picture representing the meaning of the word ("picture-association" learning). At test, either the FL elements of these paired-associate pairs are presented and the learners have to come up with the corresponding L1 word ("receptive" testing), or the L1 word (or picture) is shown and the learner is asked to produce the corresponding FL word ("productive" testing).

The data of our earlier three studies showed that translations of concrete L1 words were learned far better than those of abstract L1 words, a finding that replicated similar results of other studies (Ellis & Beaton, 1993; Lawson & Hogben, 1998). Similarly, words with a "cognate" translation in the FL (where the FL word to be learned is orthographically and phonologically similar to its L1 equivalent) were learned far better than those with a noncognate translation. This finding demonstrated the robustness of this effect, which had also been observed by others (Ellis & Beaton; Kroll, Michael, & Sankaranarayanan, 1998; but see Granger, 1993, and Meara, 1993, for a note of warning).

A third stimulus characteristic that we manipulated was the frequency of the L1 words in a paired-associate pair (de Groot & Keijzer, 2000; Lotto & de Groot, 1998). This frequency is typically highly correlated with the frequency of the corresponding FL words ($r = .78$ in de Groot, 1992), but note that this is not the case when the FL words to be learned are in fact nonwords (as is the case in the present study). In such cases, the frequency of the FL "words" is obviously zero. Unlike the variables word concreteness and cognate status, the variable frequency only caused small effects, which were not reliable across studies. Importantly, *if* an effect occurred, it was always in the same direction, with frequent L1 words showing better learning (of the corresponding FL words) than infrequent L1 words. A further finding of interest was that the words that were hardest to learn (abstract words; noncognates) were the ones that were forgotten first (de Groot & Keijzer).

The present study looked at the role of yet two further variables that might affect FL vocabulary learning. The first variable was phonotactical typicality (or simply "typicality"), a measure of the degree in which the phonological structure of the FL words to be learned resembles the sound structure of the learner's L1 (Dutch). As in de Groot and Keijzer (2000), nonwords served as the FL vocabulary to be learned (see also Hall, 2002, and Papagno, Valentine, & Baddeley, 1991). The advantage of using nonwords over existing FL words is that it is guaranteed that all vocabulary to be learned is unknown to all participants and that the "FL" materials can be manipulated systematically, here, on typicality. The second variable, music, looked at the effect of background music on learning FL vocabulary.

Typicality

The variable typicality was predicted to affect FL vocabulary learning because a substantial amount of data suggests that phonological coding of the word forms to be learned is an essential component of the learning process and that impeding such coding hinders vocabulary learning. For instance, Papagno et al. (1991) showed recall scores to be lower in a word-association learning condition (called "articulatory suppression") where learners had to utter a sound (e.g., "bla") repeatedly during learning, than in a control condition, where the learners performed a finger-tapping task while learning. Furthermore, Baddeley, Papagno, and Vallar (1988), testing a woman whose phonological memory was impaired as a consequence of having suffered a stroke, found that she was completely unable to learn nonwords that were paired with words in her L1, Italian. Both studies suggest that storage of new word forms in long-term memory is mediated by phonological coding of the new words (see Gathercole & Baddeley, 1993, and Gathercole & Thorn, 1998, for more support for this claim). Phonological coding is likely to be hampered when the learner encounters a phonotactically atypical nonword, alien to the learner. Yet, a second reason why atypical nonwords might be relatively hard to learn is that, when learning FL vocabulary, the learner exploits phonological long-term memory codes that resemble the phonology of the vocabulary to be learned (e.g., Cheung, 1996). This benefits the learning of typical nonwords, because their sound structures resemble the phonological codes stored in the learners' long-term memory. If an effect of typicality indeed materializes, it will be of interest to see whether its counterpart (as has been observed for the effects of concreteness and cognate status) also occurs: that the words hardest to learn (here, the atypical nonwords) are the ones forgotten first.

Music

Felix (1993) reviewed studies that tested whether background music affects learning. The reported studies varied on the type of music presented to the participants, the tasks to be performed (paired-associate learning being one of them), whether music was played during learning, testing, or both, and, finally, on the studies' setting: in a laboratory or in a natural teaching environment. Felix concluded that music played during learning affects performance positively, especially when the music played concerned baroque and classical pieces. She further concluded that retention is best when music is played during both learning and testing, a finding that exemplifies the well-known phenomenon of "context-dependent" memory" (i.e., that test performance is better the more similar the circumstances at test mimic those of the learning episode; e.g., Godden & Baddeley, 1975).

Although positive effects of background music occurred in the studies reviewed by Felix (1993), the pattern of results was not at all consistent. For instance, in a massive vocabulary learning study by Schuster (1985), none of a diverse set of styles of music played in the treatment conditions showed enhanced learning as compared to the no-music control conditions. Inconsistent results across studies often reflect differences in experimental procedures that are differentially sensitive to the experimental manipulations. Therefore, it seemed worthwhile to include a music manipulation in a study that adopted the experimental procedures of an earlier study that had shown robust effects of other variables. The procedures used here replicated those of de Groot and Keijzer (2000), involving word-association learning with, as already pointed out, nonwords serving as the FL words to be learned.

The music manipulation compared a treatment condition where a baroque piece was played (part of the *Brandenburg Concerto* by J. S. Bach) with a no-music (silent) control condition. The participants in the music condition only had music presented to them during learning, not at test. This resembled a common learning-then-testing experience of FL students, where they learn to pair FL words and the corresponding L1 words at home, presumably at least occasionally with music playing in the background, and are subsequently tested in a classroom setting, typically in silence.

L1 Frequency and L1 Concreteness

In addition to the variables typicality and music, the present study once again included the variable frequency (of the L1 word). As mentioned above, earlier studies (de Groot & Keijzer, 2000; Lotto & de Groot, 1998) had demonstrated small effects of this variable that had not been statistically reliable in all cases. Inclusion of this variable was meant to strengthen the (still rather weak) suggestion that it plays a role in FL word learning as well, albeit a small one. Finally, the variable L1 concreteness was included once more. Unlike the effects of L1 frequency, the effects of this variable have been shown to be large and robust. Therefore, the reason for including this variable was not so much to obtain yet further support for the role of this variable in FL vocabulary *learning*. Instead, it was included because it would provide an opportunity to also demonstrate the robustness of the L1 concreteness effect on FL vocabulary *forgetting*, a finding that to my knowledge has only been demonstrated once before (de Groot & Keijzer).

To summarize, in this study I investigated the effects of frequency and concreteness of the L1 words, typicality of the FL nonwords, and background music on FL vocabulary learning and forgetting. On the basis of earlier work, I expected a small (but possibly statistically unreliable) effect of L1 word frequency to materialize. Typicality of the FL nonwords was predicted to have an effect because of the important role that both the *process* of phonological coding and phonological *knowledge* stored in long-term memory have been shown to play in FL vocabulary learning. I expected poorer performance for atypical FL nonwords than for typical FL nonwords. As for the music manipulation, due to its more explorative nature no exact predictions were hazarded. The question to answer here was whether under circumstances that have shown to be sensitive to experimental manipulation (as demonstrated by the large and robust effects of a number of variables, especially concreteness and cognate status), an effect of the music manipulation would materialize. If not, it would be safe to conclude that such null-effect was not due to lack of sensitivity of the experimental design. Finally, I was interested to see whether our earlier finding (de Groot & Keijzer, 2000), that L1 word concreteness not only affects FL vocabulary learning but also forgetting, would be replicated in this study.

Method

Participants

Thirty-six participants took part: 18 in the music condition and 18 in the silent condition. They were all first-year psychology students at the University of Amsterdam and rather experienced FL learners. All were native speakers of Dutch who had received formal training in a number of FLs all through secondary school and whose current university training involved studying many English textbooks. Following the third session in Part 1 of the experiment (see Table 1), the participants filled in a language questionnaire that assessed their FL skills. One of the questions was to list all of the languages that the participants had any knowledge of; a second was to assess their passive (reading and listening) and active (speech and writing) language abilities in English, German, and French, the FLs most often taught in Dutch secondary schools, using a rating scale of 1 to 7 (where 7 indicated a level of proficiency equal to the corresponding skill in Dutch and 1 indicated that the skill was completely lacking). All participants listed English as a language they knew. French, German, and Spanish were also mentioned relatively often. A number of other languages were mentioned occasionally, such as Surinamese and Hindi. The average active language-ability ratings for the participants in the music condition were 5.2, 3.8, and 2.9 for English, German, and French, respectively. The corresponding ratings for the participants in the silent condition were 5.3, 3.0, and 2.9, respectively. The average passive language-ability ratings for the participants in the music condition were 6.2, 4.8, and 4.0 for English, German, and French, respectively. The corresponding ratings for the participants in the silent condition were 6.1, 3.8, and 3.7, respectively. Thus, the FL profile was similar for the participants in the two music conditions. When signing up for the experiment, the participants were informed that it would consist of two parts, a first part that would take about 2.5 hr, and a second, relatively short, part a week later. All participants received course credit for participating.

Materials

Typicality of the FL Words

The materials presented for learning were 64 stimulus pairs ("translation pairs") each consisting of a Dutch word and a nonword. The nonwords varied on phonotactical typicality:

32 were phonotactically typical in Dutch and the remaining 32 were phonotactically atypical in Dutch. The nonwords were constructed using a corpus of "positional letter frequencies" in Dutch (Rolf & Van Rijnsoever, 1984). This corpus contains all position-dependent frequencies of occurrence of each of the letters in the Dutch alphabet, as based on a corpus of 620,000 Dutch written words (Uit den Boogaart, 1975). The atypical nonwords were constructed from letters that occur infrequently on their respective letter positions in Dutch words of this particular length. For instance, "obfa" is an atypical nonword because the "o" in first position, the "b" in second position, the "f" in third position, and the "a" in fourth position all occur relatively infrequently in Dutch words containing four letters (their positional occurrence in the corpus is 3,491, 18, 86, and 386, respectively, adding up to the relatively low cumulative score of 3,981 for the nonword as a whole). In contrast, "voje" is a typical nonword because the "v" in first position, the "o" in second position, the "j" in third position, and the "e" in fourth position, are all relatively common (8,414, 17,523, 9,063, and 6,834, respectively, adding up to the high cumulative score of 41,834). All nonwords, typical and atypical, were dissimilar in form to the L1 words to which they were paired; in other words, they were all noncognates.

Imageability (=Concreteness) of the L1 Words

The Dutch words within both sets of 32 translation pairs thus formed varied on two variables: word imageability and word frequency. Imageability scores are usually derived from studies in which groups of subjects are asked to rate lists of words as to the ease or difficulty with which they evoke a mental image. The present words' scores on this variable were drawn from a corpus that the present author collected earlier (de Groot, Dannenburg, & Van Hell, 1994) and that had employed the instructions of Paivio, Yuille, and Madigan (1968) translated into Dutch, using the 7-point scale of these authors (with 1 and 7 as end points): Words that evoke images with the greatest difficulty or not at all had to be rated 1; words that arouse mental images most readily had to be assigned a 7; words intermediate in ease or difficulty of imagery had to be rated appropriately in between the two extremes (Paivio et al.). This imageability variable typically correlates highly (e.g., .83 in Paivio et al.; .96 in De Groot, 1989) with word concreteness, a variable that expresses the degree to which a word (or, rather, the entity the word refers to) can be experienced by the senses: Words easy to imagine are usually words that refer to concrete entities and words hard to imagine usually refer to abstract entities. (Typical exceptions are words with strong emotional or evaluative connotations, such as *anxiety* and *jealousy*; words for fictitious creatures, such as *demon* and *devil*; and some concrete but extremely rare words, such as *armadillo* and *encephalon*; all examples taken from Paivio et al.). Given this common confounding between imageability and concreteness, for the sake of convenience I referred to this variable as the "concreteness" variable in the introduction section to this article, and words with high and low imageability ratings were called concrete and abstract, respectively. I will continue doing so in the remainder of this article. Within each of the two sets of translation pairs–the group with the phonotactically typical nonwords and the group with the atypical nonwords–half of the L1 (Dutch) words were concrete and the remaining half were abstract.

Frequency of the L1 Words

The L1 words' frequency values were taken from the CELEX corpus, which is based on a count of 42.5 million printed Dutch words (Baayen, Piepenbrock, & Van Rijn, 1993). The values used here were the logarithms of the total number of occurrences of the selected L1 words in this corpus. Half of the concrete words within each of the two sets of translation pairs were relatively frequent, whereas the remaining half were infrequent. Similarly, half of the abstract words within each of the two sets of translation pairs were relatively frequent, whereas the remaining half of the abstract words were infrequent.

Further constraints in selecting the materials were (a) that frequent and infrequent L1 words were matched on L1 concreteness, (b) that concrete and abstract L1 words were matched on L1 frequency, and (c) that the translation pairs with typical and atypical nonwords were matched on L1 concreteness and L1 frequency as well as on the length of the L1 words in these pairs. A set of t-tests showed that these criteria were met ($p > .10$ in all cases). Finally, all words and nonwords consisted of four, five, or six letters.

The appendix lists the score on each of the above-listed three variables for each of the 64 translation pairs. It also shows the means and corresponding standard deviations for the eight groups of translation pairs formed by the orthogonal manipulation of the three variables.

Apparatus and Procedure

The experiment was run on an Apple Macintosh computer. The stimuli were presented in black lowercase letters on a light gray background in the center of the screen. A Pascal program controlled the stimulus presentation during learning and testing. All participants were tested individually in a room that was dimly lit. The study was split up in two parts: Part 1 consisted of three learning-then-test sessions; Part 2 concerned a retest that was held 1 week later. Table 1 provides a summary of the complete study.

Learning-Then-Test Sessions

Each of the three learning-then-test sessions consisted of *two* learning rounds, in each of which all of the 64 translation pairs were presented once, followed by *one* test phase. On each trial of a learning round, a translation pair (a Dutch word and its “translation”; i.e., the nonword to which the word was paired) appeared on the screen for 10 s, preceded by a fixation stimulus (an asterisk). The Dutch word and its translation appeared

Table 1

Summary of the study

Part 1, Session 1	
	First presentation of all L1-FL pairs for *learning*
	Second presentation of all L1-FL pairs for *learning*
	First *testing* (Test 1)
Part 1, Session 2	
	Third presentation of all L1-FL pairs for *learning*
	Fourth presentation of all L1-FL pairs for *learning*
	Second *testing* (Test 2)
Part 1, Session 3	
	Fifth presentation of all L1-FL pairs for *learning*
	Sixth presentation of all L1-FL pairs for *learning*
	Third *testing* (Test 3)
Part 2, Retest (1 week later)	
	Fourth *testing* (Test 4)

simultaneously, next to one another (the Dutch word left of its translation). The 64 translation pairs of a round were presented in a random order, a different order for each participant. After the presentation of 32 of the translation pairs of a round, the participant was allowed a brief rest before the remaining 32 pairs appeared. The participants did not receive any instructions as to what learning strategy to adopt. Immediately following the complete set of 64 translation pairs, the complete set was presented for learning a second time, using exactly the same procedure as during the first round. Eighteen of the participants learned in silence, and the remaining 18 learned while part of the fourth *Brandenburg Concerto* by J. S. Bach played in the background, the same (12 min) part in both learning rounds (and in the remaining four learning rounds yet to follow). The piece can be characterized as classical, baroque, instrumental, and rhythmical. It involved two flutes and a violin set against an ensemble of string instruments.

After completion of the second learning round, the first test (Test 1) started. During this test, the 64 newly learned FL words (the nonwords) were presented on the screen for translation into Dutch, one by one and in random order. Prior to each FL word, a fixation stimulus (an asterisk) appeared in the middle of the screen for 1 s. The moment this fixation stimulus disappeared, the test stimulus appeared and remained on the screen until the response was given. The experimenter, sitting next to the participant, typed the participant's response (what was being typed did not appear on the screen) and then started the next trial by touching the return key. The maximum presentation duration of the test stimulus was 10 s. If the participant had not provided a response within 10 s, the test stimulus disappeared from the screen and a new one was presented, again preceded by a fixation stimulus. A tape recorder recorded the participant's responses, so that responses missed by the experimenter could be identified at a later point. During testing, neither group of participants had music played to them in the background.

Two more such sessions of two learning rounds followed by one test immediately followed the first session. In all, during Part 1 of the study there were six learning occasions per translation pair and the amount of learning was tested three times. After completing Part 1, the participant filled in the questionnaire and an appointment for the retest a week later was made. The participants were not explicitly informed what they would be asked to do on their next visit to the laboratory.

Retest

During the retest, the participants were tested once more without first relearning the translation pairs. The circumstances of this retest were identical to those of the tests 1 week earlier.

Results

For the moment ignoring the concreteness variable, for each of the 36 participants, 16 recall scores were calculated, namely one for each of the 2 (typicality: typical vs. atypical FL words) × 2 (frequency: frequent vs. infrequent L1 words) × 4 (test) conditions. Each score reflected the percentage of correct responses of that particular participant in that particular Typicality × Frequency × Test condition. These scores served as input for the analyses by participants (see the next paragraph). Furthermore, for both the music condition and the silent condition, a recall score was calculated for each of the 64 items in each of the four recall tests. Each of these scores reflected the percentage of participants who had correctly recalled the response to that particular stimulus word in that particular Test × Music condition. These scores served as input for the analyses by items (see the next paragraph). The analyses by participants were meant to indicate whether a particular effect generalizes over participants; the analyses by items—less common in FL vocabulary learning research—were meant to indicate whether a particular effect generalizes over items.

On the recall scores thus calculated, two pairs of analyses were performed, each pair consisting of one analysis by participants and a second by items. The goal of one pair of analyses was to chart the learning process across the three learning-then-test sessions. The data included in these analyses concerned the recall scores for the three tests in Part 1 of the study (Test 1, Test 2, and Test 3; see Table 1). The goal of the second pair of analyses was to reveal how much of the newly learned FL vocabulary was forgotten 1 week after learning. These analyses were based on the recall scores of Test 3 in Part 1 of the study and of Test 4 in Part 2 of the study.

As pointed out in the introduction to this article, the concreteness manipulation was included to see whether not only the (robust) effect of this variable on learning but also its effect on forgetting materialized again. To assess the effect of this variable, a set of additional analyses was performed on the data. These were the same as those described earlier, except that they included concreteness as an additional variable. To save space and maximize readability, the results of these analyses will not be presented in detail. Instead, the exclusive focus will be on whether these analyses replicate the main findings of de Groot and Keijzer (2000), namely that FL translations of concrete L1 words are easier to learn and, especially, less easily forgotten than those of abstract words.

Learning

The analyses performed to chart the learning process concerned two 2 (frequency) $\times$ 2 (typicality) $\times$ 2 (music) $\times$ 3 (test: Test 1 vs. Test 2 vs. Test 3) analyses of variance (ANOVAs), one by participants and a second by items. In the analysis by participants, the variables frequency, typicality, and test were treated as within-participants variables, whereas the variable music was treated as a between-participants variable. In the analysis by items, the variables test and music were treated as within-items variables, whereas the variables frequency and typicality were treated as between-items variables. An α-level of $p < .05$ was adopted for all statistical tests. In the presentation of the results, $F1$ refers to the analysis by participants and $F2$ to the analysis by items. Recall scores will be presented in percentages. The recall scores for all 24 cells formed by the variables frequency, typicality, test, and music are presented in Table 2 (in the rows for Test 1, Test 2, and Test 3) and in Figure 1.

In all cases, the recall scores for the translation pairs with atypical nonwords were lower than those for the pairs with typical nonwords in the comparable cell in the experiment (e.g., the cell "Test 1, music, typical, frequent" compared with the cell "Test 1, music, atypical, frequent"). Similarly, in all cases, the recall scores were lower for a particular cell in the silent condition than for the analogous cell in the music condition (e.g., the cell, "Test 1, music, typical, frequent" compared with the cell "Test 1, silence, typical, frequent"). Figure 1 visualizes this beneficial effect of music.

The main effects of typicality and test were significant both by participants and by items: typicality: $F1(1, 34) = 91.56$, $p < .001$, $\eta^2 = .73$, and $F2(1, 60) = 21.63$, $p < .001$, $\eta^2 = .27$ (in fact, here and elsewhere, η^2 refers to *partial* eta squared);

Table 2

Recall scores (in percentages) for all Frequency × Typicality × Test × Music conditions

	Typical frequent	Typical infrequent	Atypical frequent	Atypical infrequent
Music				
Test 1	48.6 (26.4)	40.3 (26.1)	23.3 (18.9)	25.7 (26.1)
Test 2	86.1 (16.8)	83.0 (15.3)	68.8 (22.3)	66.0 (24.6)
Test 3	91.7 (9.4)	95.1 (9.1)	86.1 (14.3)	89.2 (14.7)
Test 4	69.8 (22.6)	54.2 (21.5)	50.0 (24.3)	47.2 (23.7)
Silence				
Test 1	47.9 (30.1)	30.2 (25.4)	22.2 (19.6)	19.1 (18.5)
Test 2	72.9 (32.0)	66.0 (32.8)	61.8 (31.3)	56.9 (31.3)
Test 3	84.7 (24.7)	86.5 (19.9)	79.9 (20.1)	71.5 (28.6)
Test 4	62.5 (25.9)	47.6 (27.7)	43.1 (24.9)	43.1 (29.5)

Note. Test 4 concerns the retest one week after learning. Frequency concerns the L1 words in the translation pairs. Typicality concerns the FL words in the translation pairs. Standard deviations are in parentheses.

test: $F1(2, 68) = 538.73$, $p < .001$, $\eta^2 = .94$, and $F2(2, 120) = 1098.49$, $p < .001$, $\eta^2 = .95$. The recall score was 13.5% larger for typical nonwords than for atypical nonwords (69.4% vs. 55.9%); the recall score increased from 32.2% in Test 1, to 70.2% in Test 2, to 85.6% in Test 3. The main effect of music was statistically significant in the analysis by items, $F2(1, 60) = 91.23$, $p < .001$, $\eta^2 = .60$, but not in the analysis by participants, $F1(1, 34) = 2.55$, $p > .10$, $\eta^2 = .07$. The recall score was 8.7% larger in the music condition than in the silent condition (67.0% vs. 58.3%). Conversely, the main effect of frequency was statistically significant in the analysis by participants, $F1(1, 34) = 12.52$, $p < .01$, $\eta^2 = .27$, but not in the analysis by items, $F2(1, 60) = 1.62$, $p > .10$, $\eta^2 = .03$. The recall score for frequent words was 3.7% larger than for infrequent words (64.5% vs. 60.8%).

The two-way interaction between music and frequency was statistically significant by participants, $F1(1, 34) = 7.34$, $p = .01$,

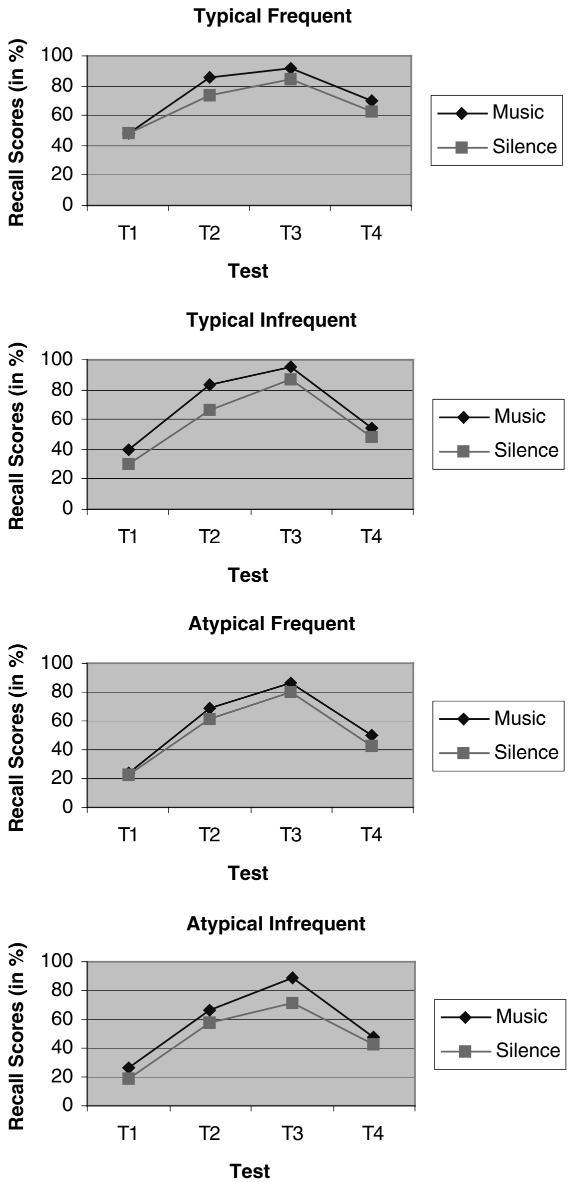

Figure 1. Recall scores (in percentages) for all Frequency × Typicality × Test (T) × Music conditions. T4 concerns the retest 1 week after learning. Frequency concerns the L1 words in the translation pairs. Typicality concerns the FL words in the translation pairs.

$\eta^2 = .18$, as well as by items, $F1(1, 60) = 9.74, p < .01, \eta^2 = .14$. The cell means of this interaction showed that the effect of the music manipulation was larger for infrequent words (66.6% vs. 55.0%) than for frequent words (67.4% vs. 61.6%); infrequent words thus benefited from the presence of background music by 11.6%, whereas frequent words only benefited from music by 5.8%.

The two-way interactions between test on the one hand and the variables frequency and typicality on the other hand were statistically significant, both by participants and by items: Test × Frequency, $F1(2, 68) = 5.60, p < .01, \eta^2 = .14$, and $F2(2, 120) = 4.20, p < .05, \eta^2 = .07$; Test × Typicality, $F1(2, 68) = 8.29, p < .01, \eta^2 = .20$, and $F2(2, 120) = 11.74, p < .001, \eta^2 = .16$. The two-way interaction between test and music was significant by items, but only marginally so by participants: $F1(2, 68) = 2.34, p = .10, \eta^2 = .07$, and $F2(2, 120) = 7.42, p < .01, \eta^2 = .11$. The cell means for these interactions are shown in Table 3. As can be seen, the effects of frequency and typicality were larger in the earlier stages of learning than in the latest stage, Test 3, and they were largest in the earliest stage, Test 1. In fact, in Test 3, the infrequent words have caught up with the frequent words. In contrast, the effect of music was considerably smaller in Test 1 than in Tests 2 and 3. None of the remaining two-way interactions were statistically significant.

The three-way interaction among typicality, frequency, and test was significant both by participants and by items: $F1(2, 68) = 9.40, p < .001, \eta^2 = .22$, and $F2(2, 120) = 7.45, p < .01, \eta^2 = .11$. This analysis qualified the two-way interactions between test and frequency and test and typicality by showing that the larger effect of frequency in the earlier stages of learning only held for the typical nonwords in Test 1. The frequency effect was 13.1% in Condition Typical-Test 1, whereas it was rather similar in the five remaining Typicality × Test conditions (varying between –2.6% and 5.0%). Finally, the three-way interaction among music, test, and typicality was significant by items, $F2(2, 120) = 4.49, p < .05, \eta^2 = .07$, but not by participants, $F1(2, 68) = 2.04$,

Table 3

Learning analyses; recall scores (in percentages) for the interactions between test and frequency, test and typicality, test and concreteness, and test and music

	Test × Frequency		
	Frequent	Infrequent	Frequency effect
Test 1	35.5 (27.1)	28.8 (23.9)	6.7
Test 2	72.4 (27.5)	68.0 (28.1)	4.4
Test 3	85.6 (18.3)	85.6 (21.1)	0.0
	Test × Typicality		
	Typical	Atypical	Typicality effect
Test 1	41.8 (27.8)	22.6 (19.4)	19.2
Test 2	77.0 (26.6)	63.4 (27.7)	13.6
Test 3	89.5 (17.5)	81.7 (21.2)	7.8
	Test × Concreteness		
	Concrete	Abstract	Concreteness effect
Test 1	40.3 (27.0)	24.0 (21.7)	16.3
Test 2	78.4 (24.7)	62.0 (28.6)	16.4
Test 3	89.4 (15.7)	81.8 (22.5)	7.6
	Test × Music		
	Music	Silence	Music effect
Test 1	34.5 (25.3)	29.9 (26.2)	4.6
Test 2	76.0 (21.7)	64.4 (32.1)	11.6
Test 3	90.5 (12.5)	80.6 (24.1)	9.9

Note. Frequency and concreteness concern the L1 words in the translation pairs. Typicality concerns the FL words in the translation pairs. Standard deviations are in parentheses. The interactions between test and frequency, test and typicality, and test and concreteness are significant both by participants and by items. The interaction between test and music is marginally significant by participants and significant by items.

$p > .10$, $\eta^2 = .06$. The interaction appeared to be due to the fact that in the condition with typical FL words, the effect of music was larger in Test 2 (15.1%) than in Test 3 (7.8%), whereas the opposite pattern occurred in the condition with atypical FL words (Test 2: 8.0%; Test 3: 12.0%). In Test 1, the effect of music was about equally large in both typicality conditions (typical: 5.3%; atypical: 3.8%). This result, for which no obvious interpretation comes to mind, does not affect the above conclusion that the effect of music appeared to be smaller in Test 1 than in the later test phases; it seems to hold for both typical and atypical words. None of the remaining interactions was significant on either analysis.

Forgetting

The analyses performed to see how much forgetting had occurred during the 1-week interval between learning and retest, concerned two 2 (frequency) $\times$ 2 (typicality) $\times$ 2 (music) $\times$ 2 (test: Test 3 vs. Test 4) ANOVAs, again one by participants and a second by items. The recall scores for all 16 cells formed by the four variables are presented in Table 2 (in the rows for Tests 3 and 4; see also Figure 1, T3 and T4). As shown in Table 2, 1 week after learning, recall for all types of translation pairs was better when during learning music had been played to the participants than when they had learned in silence (e.g., compare cell "T4, music, typical, frequent" with cell "T4, silence, typical, frequent").

In the forgetting analyses, the interactions between test on the one hand and the remaining variables on the other hand are particularly important, because these, and *not* the main effects, point at the amount of forgetting. However, for the sake of consistency with the learning analyses, the main effects will be reported as well.

The main effects of typicality and test were significant both by participants and by items: typicality: $F1(1, 34) = 47.31$, $p < .001$, $\eta^2 = .58$, and $F2(1, 60) = 12.94$, $p < .01$, $\eta^2 = .18$; test:

$F1(1, 34) = 226.56$, $p < .001$, $\eta^2 = .87$, and $F2(1, 60) = 540.77$, $p < .001$, $\eta^2 = .90$. The recall score was 10.2% larger for typical nonwords than for atypical nonwords (74.0% vs. 63.8%); the recall score decreased from 85.6% in Test 3 to 52.2% in Test 4, a decrease of 33.4%. The main effect of music was statistically significant in the analysis by items, $F2(1, 60) = 39.28$, $p < .001$, $\eta^2 = .40$, but not in the analysis by participants, $F1(1, 34) = 2.73$, $p > .10$, $\eta^2 = .07$. The recall score was 8.1% larger in the music condition than in the silent condition (72.9% vs. 64.8%). Conversely, the main effect of frequency was statistically significant in the analysis by participants, $F1(1, 34) = 10.56$, $p < .01$, $\eta^2 = .24$, but not in the analysis by items, $F2(1, 60) = 2.14$, $p > .10$, $\eta^2 = .03$. The recall score for frequent words was 4.2% larger than for infrequent words (71.0% vs. 66.8%).

The (particularly relevant; see earlier in this section) two-way interactions between test on the one hand and the variables frequency and typicality on the other hand were both statistically significant by participants; on the corresponding item analysis, the interaction between test and frequency was statistically significant, but the one between test and typicality only approached significance: Test × Frequency, $F1(1, 34) = 21.87$, $p < .001$, $\eta^2 = .39$, and $F2(1, 60) = 8.41$, $p < .01$, $\eta^2 = .12$; Test × Typicality, $F1(1, 34) = 5.16$, $p < .05$, $\eta^2 = .13$, and $F2(1, 60) = 2.87$, $p = .10$, $\eta^2 = .05$. The cell means for these interactions are shown in Table 4. More forgetting had occurred for the materials that were relatively hard to learn, that is, for infrequent words and atypical nonwords, than for frequent words and typical nonwords, respectively. None of the remaining two-way interactions was statistically significant on either the analysis by participants or the analysis by items.

Of the three-way interactions, the one among typicality, frequency, and test was significant both by participants and by items: $F1(1, 34) = 18.74$, $p < .001$, $\eta^2 = .36$, and $F2(1, 60) = 11.04$, $p < .01$, $\eta^2 = .16$, thus qualifying the interactions between test and frequency and test and typicality. The cell means of this interaction are presented in the upper part of Table 5. As shown there,

Table 4

Forgetting analyses; recall scores (in percentages) for the interactions between test and frequency, test and typicality, and test and concreteness

	Test 3	Test 4	Amount of forgetting
Test × Frequency			
Frequent	85.6 (18.3)	56.3 (26.4)	29.3
Infrequent	85.6 (21.1)	48.0 (25.8)	37.6
Test × Typicality			
Typical	89.5 (17.5)	58.5 (25.7)	31.0
Atypical	81.7 (21.2)	45.8 (25.6)	35.9
Test × Concreteness			
Concrete	89.4 (15.7)	62.7 (24.6)	26.7
Abstract	81.8 (22.5)	41.7 (23.8)	40.1

Note. Frequency and concreteness concern the L1 words in the translation pairs. Typicality concerns the FL words in the translation pairs. Test 3 concerns testing after the fifth and sixth learning rounds. Test 4 concerns the retest 1 week after learning. Standard deviations are in parentheses. The interactions between test and frequency and between test and concreteness are significant both by participants and by items. The interaction between test and typicality is significant by participants and marginally significant by items.

relatively little forgetting had occurred for typical FL nonwords paired with frequent L1 words (22.1%), whereas the amount of forgetting was about equally large in the three remaining Typicality × Frequency conditions (between 35.3% and 39.9%). In other words, the least forgetting had occurred for the words that had been the easiest to learn. (Collapsing across both the two levels of the variable music and the three levels of the variable test in the learning part of the experiment [see Table 2] shows that the conditions typical/frequent, typical/infrequent, atypical/frequent, and atypical/infrequent produced overall learning scores of 72.0%, 66.8%, 57.0%, and 54.7%, respectively.)

Finally, the three-way interaction among music, test, and frequency was significant, $F1(1, 34) = 5.47, p < .05, \eta^2 = .14$, and $F2(1, 60) = 4.95, p < .05, \eta^2 = .08$. The cell means of this interaction are presented in the lower part of Table 5. The interaction

Table 5

Forgetting analyses; recall scores (in percentages) for the interactions among typicality, frequency, and test (upper part) and among music, frequency, and test (lower part)

	Frequent			Infrequent		
	Test 3	Test 4	Forgetting	Test 3	Test 4	Forgetting
Typicality × Frequency × Test						
Typical	88.2 (18.9)	66.1 (24.4)	22.1	90.8 (16.0)	50.9 (24.9)	39.9
Atypical	83.0 (17.6)	46.5 (24.7)	36.5	80.4 (24.3)	45.1 (26.7)	35.3
Music × Frequency × Test						
Music	88.9 (12.3)	59.9 (25.3)	29.0	92.2 (12.5)	50.7 (22.8)	41.5
Silence	82.3 (22.5)	52.8 (27.0)	29.5	79.0 (25.6)	45.3 (28.5)	33.7

Note. Frequency concerns the L1 words in the translation pairs. Typicality concerns the FL words in the translation pairs. Test 3 concerns testing after the fifth and sixth learning rounds. Test 4 concerns the retest 1 week after learning. Standard deviations are in parentheses. Both interactions are significant both by participants and by items.

showed that in condition infrequent/music more forgetting had occurred (41.5%) than in any of the remaining three Frequency × Music conditions (between 29.0% and 33.7%). The learning analyses had shown that infrequent words benefited more from background music than frequent words. The present interaction suggests that this extra profit is not a lasting effect. None of the remaining interactions was significant on either analysis.

Effects of Concreteness

On the analogous set of learning analyses that included concreteness as an additional variable, the main effect of concreteness and the interaction between concreteness and test were both significant by participants and by items: main effect, $F1(1, 34) = 136.23$, $p < .001$, $\eta^2 = .80$, and $F2(1, 56) = 31.84$, $p < .001$, $\eta^2 = .36$; interaction, $F1(2, 68) = 9.83$, $p < .001$, $\eta^2 = .22$, and $F2(2, 112) = 11.02$, $p < .001$, $\eta^2 = .16$. Collapsed across the three tests of Part 1 of the experiment, the recall score was 13.5% larger for concrete words than for abstract words (69.4% vs. 55.9%). The interaction data are presented in Table 3. Similarly to the effects of frequency and typicality, the concreteness effect was especially large in the earlier stages of learning. The corresponding set of forgetting analyses showed an interaction between concreteness and test: $F1(1, 34) = 36.41$, $p < .001$, $\eta^2 = .52$, and $F2(1, 56) = 33.31$, $p < .001$, $\eta^2 = .37$. This interaction, shown in Table 4, demonstrated that over the 1-week interval between learning and retest, more forgetting had occurred for abstract words than for concrete words. For reasons explained earlier, no further details of this set of analyses will be presented.

Table 6 summarizes the major findings of this study.

Discussion

Effects of the Stimulus Characteristics

L1 Word Concreteness

The present investigation replicated the effect of L1 word concreteness as observed in earlier studies (de Groot & Keijzer,

Table 6

Summary of the major findings

- More FL words with an L1-typical than with an L1-atypical form were learned. This typicality effect was especially large in the earlier stages of learning.
- More FL words were learned when paired with frequent L1 words than when paired with infrequent L1 words. This frequency effect was especially large in the earlier stages of learning.
- More FL words were learned when paired with concrete L1 words than when paired with abstract L1 words. This concreteness effect was especially large in the earlier stages of learning.
- More FL words were learned in the music condition than in the silent condition. This music effect was especially large in the later stages of learning (but did not generalize over participants).
- Typical FL words were less susceptible to forgetting than atypical FL words.
- FL words paired with frequent L1 words during learning were less susceptible to forgetting than those paired with infrequent L1 words.
- FL words paired with concrete L1 words during learning were less susceptible to forgetting than those paired with abstract L1 words.
- In general, the FL words that were learned best were the least susceptible to forgetting.

2000; Ellis & Beaton, 1993; Lawson & Hogben, 1998; Van Hell & Candia Mahn, 1997): Recall was higher for FL (non)words paired with concrete L1 words than for those paired with abstract L1 words. Collapsed across the three tests in the learning part of the study, the concreteness effect was 13.5%. Furthermore, as in de Groot and Keijzer, the retest 1 week after training showed that more forgetting had occurred for FL nonwords that, during learning, had been paired with abstract L1 words (40.1% forgetting) than for those paired with concrete L1 words (26.7% forgetting; see Table 4).

L1 Word Frequency

First language word frequency showed the same pattern of results as obtained by de Groot and Keijzer (2000, Experiment 2b): It was statistically significant on the analyses by participants but not on the analyses by items, indicating that the effect was attributable to a subset of the stimuli only. It deviated from the results obtained by Lotto and de Groot (1998), where the effect was reliable both by participants and by items. However, across all of these studies, recall scores of translation pairs with frequent L1 words were larger than for those with infrequent L1 words, although the effect was much smaller than the concreteness effect. Collapsed across the three tests in the learning part of the present study, the L1 frequency effect was 3.7%. The interaction with the variable test showed that the effect of L1 frequency gradually decreased over the three tests and ceased to exist at the third test (i.e., after six learning occurrences; Table 3). In other words, the learning of FL nonwords paired with infrequent L1 words gradually caught up with those paired with frequent L1 words. The forgetting analyses showed that 1 week after learning, more FL forms paired with infrequent L1 words had been lost (37.6%) than FL forms paired with frequent L1 words (29.3%; Table 4). Unlike the main effect of frequency in the learning analyses, this differential forgetting for translation pairs with frequent and infrequent L1 words was reliable both by participants and by items.

FL Typicality

The new variable, FL nonword typicality, showed a pattern resembling the effects of L1 concreteness and frequency. The typical nonwords were learned better than the atypical nonwords. Collapsed across the three tests in the learning part of the study, the typicality effect was 13.5%—incidentally exactly the same as the concreteness effect (and much larger than the L1 frequency effect). The interaction with the variable test showed that the effect of FL typicality decreased over the three tests, performance on trials with atypical nonwords gradually catching up with performance for typical nonwords (Table 3). The forgetting analyses showed that at retest 1 week after learning, (slightly) more forgetting had occurred for translation pairs with atypical nonwords

(35.9%) than for those with typical nonwords (31.0%; Table 4). However, this differential forgetting for typical and atypical nonwords was only marginally significant by items, indicating that the effect was attributable to a subset of the stimuli only.

Overall, this combination of results once again suggests that stimuli that are relatively easy to learn are less susceptible to forgetting than stimuli that are relatively hard to learn. Whereas de Groot and Keijzer (2000) showed this effect while manipulating characteristics of the L1 words in the translation pairs (or, in the case of cognate status, the similarity between the L1 word and its FL translation), the present study shows that it also holds if the FL word in the translation pairs is manipulated. A final finding worth mentioning here is the interaction between the variables frequency, typicality, and test in the forgetting analyses. It suggested that of the four groups of learning material formed by the orthogonal manipulation of the variables frequency and typicality, the frequent L1 words paired with typical FL nonwords are least susceptible to forgetting (Table 5, upper part). This result is compatible with the general pattern in the data that the material easiest to learn has the best chance of being stored in memory in a lasting way.

The Effects of the Stimulus Characteristics Explained

L1 Word Concreteness

De Groot and Keijzer (2000) explained the concreteness effect in terms of differences in information density between the memory representations of concrete and abstract L1 words, with the representations of concrete words containing more information than those of abstract words: Concrete words are assumed to both have an image representation in memory as well as a verbal representation (Paivio, 1986; Paivio & Desrochers, 1980), whereas for abstract words only a verbal representation is assumed to be stored. Alternatively, the memory representations of concrete words are thought to contain more information elements than those of abstract words in an "amodal" memory system (that stores all input in a form that is neutral to the perceptual characteristics of the input and, consequently, does not distinguish between image and verbal representations; de Groot, 1989; Kieras, 1978). The denser representations of concrete L1 words enable more opportunities for anchoring the FL word's name onto the memory representation of the corresponding L1 word. This might both explain the better learning performance for FL translations of concrete L1 words as well as the finding that the FL translations of concrete L1 words are less susceptible to forgetting than those of abstract L1 words.

L1 Word Frequency

There is some evidence that the small frequency effect can be explained in the same way: In a "continued word association" study (where participants are presented with a set of words and are asked to come up with all the words that they think of when reading each word in the set), de Groot (1989) showed that frequent words produced slightly more associations within a given unit time than infrequent words, suggesting that the memory representations of frequent words contain slightly more information elements. Consequently, frequent L1 words enable more opportunities to anchor new information (here, the corresponding FL word's names) onto their memory representations. Word frequency is highly correlated with concept familiarity: Words that the language user comes across relatively often (in print—the frequency measure employed here—or in spoken language) likely refer to concepts that are more familiar to the language user than concepts for words that the language user does not encounter often (in fact, word familiarity is usually determined by asking participants to indicate how often they have seen, heard, or used the word; see, e.g., de Groot, 1992, and Paivio, 1968, for correlations between printed word frequency and familiarity determined this way). Therefore, instead of L1 (written) word frequency, concept familiarity might ultimately underlie the observed effects of frequency. Importantly, differences in concept familiarity are plausibly reflected in differential informational density of the memory representations, familiar concepts having denser representations than unfamiliar concepts. Ultimately then, the effect can again be attributed to differences between the memory representations for frequent (familiar) and infrequent (unfamiliar) words.

If this account of the frequency and concreteness effects is correct, matching frequent and infrequent L1 words, or concrete and abstract L1 words, on the number of information elements in their memory representations should make the frequency and concreteness effects in FL vocabulary learning disappear. Consistent with this hypothesis, Sjarbaini (1998) demonstrated that the concreteness effect disappeared when concrete and abstract words were matched on "context availability," a measure that presumably reflects information density in the stored memory representations. Similarly, de Groot and Keijzer (2000) argued that such matching (but in this case incidental) might have been the reason why the frequency effect observed in their study did not generalize over items; it might also explain this same result in the present study.

FL Typicality

As hypothesized in the introduction to this article, the reason that phonotactically atypical nonwords are harder to learn and more susceptible to forgetting than typical nonwords presumably is that phonological coding is hampered when the learner encounters a phonotactically atypical nonword, alien to the learner. Such phonological coding was argued to play a crucial role in transferring the vocabulary presented for learning into long-term memory (e.g., Baddeley et al., 1988; Papagno et al., 1991). Additionally, only the sound structures of typical nonwords, not those of atypical ones, resemble the phonological structures already stored in the learners' memory prior to the onset of learning. This was hypothesized to be a second reason why atypical nonwords are relatively hard to learn (e.g., Cheung, 1996). Note that the latter interpretation is reminiscent of the interpretation of the effect of L1 concreteness (and of L1 frequency, if it occurs at all) in terms of L1 information already stored in memory prior to the FL learning episode and the anchoring opportunities provided by this stored information: The stored phonological knowledge regarding L1 (similar to the phonology of the typical nonwords but not the atypical nonwords) benefits the storage of typical nonwords. As pointed out in the introduction to this article, similarity between the L1 and the FL to be learned is one of the factors that determine FL vocabulary acquisition (e.g., Ard & Homburg, 1992). The present study suggests that phonotactical similarity might (partly) underlie the effect of language similarity: Learning typical nonwords might experimentally implement vocabulary learning in an FL related to the L1, whereas learning atypical nonwords might be the experimental analogue of vocabulary learning in a language unrelated to the L1.

The finding that the newly learned typical nonwords are (slightly) less susceptible to forgetting than the atypical nonwords extends the analogous results obtained with the variables concreteness and cognate status (de Groot & Keijzer, 2000). This finding once more suggests that words relatively hard to learn are also relatively easy to forget. Schneider, Healy, and Bourne (2002) noted that, at first sight, this finding seemed inconsistent with the results of their study, results that indicated that the more difficult the conditions under which learning takes place, the better long-term retention of newly acquired FL vocabulary is. At the same time, recall immediately after learning is better the easier the learning conditions. This pattern emerged across three manipulations of task difficulty: blocking vocabulary items by category during learning or presenting them mixed; receptive or productive learning and immediate testing; pretraining or not pretraining the participants on the new FL vocabulary to be learned (Schneider et al., 2002). From these results, they concluded that “any manipulation that increases the difficulty of the learning task may have different effects on initial and eventual performance" and that "variables that optimize training are not necessarily optimal for retention" (p. 439). Additional support for this conclusion comes from studies that have compared the effectiveness of keyword mnemonics and rote learning: Whereas quite a few studies (e.g., Hogben & Lawson, 1997; Lawson & Hogben, 1998) have shown that keyword mnemonics leads to superior recall on an immediate test (although perhaps not for all learner groups; McDaniel & Pressley, 1984; Van Hell & Candia Mahn, 1997), delayed testing following keyword mnemonics can produce lower retention scores than following rote learning (Wang & Thomas, 1999; but see Lawson & Hogben).

As acknowledged by Schneider et al. (2002), the contradiction between their study and de Groot and Keijzer's (2000) study (and the present one) can be resolved by distinguishing between the difficulty of the *learning procedures* and the difficulty of the *materials to be learned*. It is the latter, but not the former, that was manipulated by de Groot and Keijzer (2000) and in the present study. Combining the results of the various studies, we can conclude that it appears that "easy" words learned under difficult conditions have the highest chance of leaving a permanent trace in long-term memory following learning. Another conclusion to be drawn is that the newly learned FL words are not all stored in equally stable memory representations immediately after learning. This finding fits the views of Atkinson (1972), who distinguished between "permanent" and "temporary" memory states of the learned elements immediately after training. In these terms, immediately after training, the translations of concrete L1 words, of frequent L1 words, and of cognate L1 words (see the introduction to this article) have reached a permanent memory state relatively often (as compared to the translations of abstract, infrequent, and noncognate L1 words, respectively); similarly, immediately after training, typical FL words appear to enjoy a permanent state (slightly) more often than atypical FL words.

The Effect of Background Music

The results of the music variable (an environmental variable) deviated from the results of the variables that manipulated characteristics of either the L1 word or its paired FL word in two ways: (a) Whereas the effects of L1 concreteness, L1 frequency, and FL typicality in the learning analyses were relatively large immediately after the first learning-then-test round and smallest after the third, the effect of music was smallest at the first test; (b) whereas all three stimulus variables had shown differential forgetting, with relatively many of the FL words hardest to learn forgotten 1 week after learning, the statistical tests suggested that the music and silent conditions led to equal amounts of forgetting over the 1-week test-retest interval. Other noteworthy findings were (c) that the main effect of the music manipulation on learning only generalized over items, not over participants, suggesting that only a subgroup of the participants in the music condition benefited from the presence of background music, and (d) that the effect of music in the learning analyses was greater for infrequent words than for frequent words. Background music increased the learning of infrequent words by 11.6%, whereas it boosted the learning of frequent words by 5.8% only (see the Results section). This latter finding generalized over both participants and items. However, the forgetting analyses, and in particular the interaction among music, frequency, and test (see Table 5), suggested that this differential effect of music on frequent and infrequent words does not last. Importantly, a negative effect of the presence of background music never occurred.

Excepting one, I will not hazard an explanation as to why the above effects of the music manipulation might occur, as they would all be highly speculative. The exception is the finding that only a subset of the participants benefited from the presence of background music (Result c); it is excepted because of the existence of experimental support for the hypothesis that individual differences in personality—and, ultimately, individual differences in neurological thresholds of arousal in the brain—underlie this effect (Daoussis & McKelvie, 1986; Furnham & Allass, 1999; Furnham & Bradley, 1997).

Furnham and Allass manipulated the personality trait extraversion and observed that introverts performed substantially better in a silent condition than in two (pop) music conditions (simple and complex) while executing an observation task, an immediate recall task, and a delayed recall task, whereas extraverts generally performed better in the music condition. The detrimental effect of music for the introverts was greater in the complex condition than in the simple condition. Again, this pattern reversed for extraverts. As in the present study (in the analysis by participants), the main effect of music was not statistically significant. As Furnham and Allass (1999), Furnham and Bradley (1997) also demonstrated an interaction between the introvert/extravert variable and the music manipulation on two cognitive tests, one testing reading comprehension and the second testing memory, and Daoussis and McKelvie showed a similar interaction in a study testing reading comprehension. The results of the latter two studies differed from those of Furnham and Allass in that music had a detrimental effect on the cognitive performance of introverts, whereas the performance of the extraverts was relatively immune to the music manipulation (rather than being facilitated by music). However, all three studies converged on one and the same conclusion, namely that the introvert/extravert personality trait plays a role in the effects of background music on cognitive performance.

The authors of all three studies predicted the interaction between this personality trait and background music from Eysenck's (1967) theory on differential cortical arousal levels of introverts and extraverts. According to this theory, the neurological threshold of arousal is lower in introverts than in extraverts. As a consequence, optimal performance in introverts occurs at relatively low levels of stimulation, whereas optimal performance in extraverts occurs at relatively high levels of stimulation (Furnham & Allass). Presumably without being aware of this alleged underlying physiological cause, introverts and extraverts are apparently aware of the effect of background music on their study success, because extraverts claim to play background music more often while studying than introverts (Daoussis & McKelvie; Furnham & Bradley).

This account of the effects of background music on performance in cognitive tasks provides a plausible explanation of the present finding that the effect of background music did not generalize over all participants. In the present study, the personality trait extraversion was not taken into account (unfortunately, in retrospect) and the participant sample most likely included both introverts and extraverts. Beneficial effects of background music on the extraverts might have been canceled out partly by detrimental effects on the introverts. In addition to extraversion (and, ultimately, introverts' and extraverts' neurological thresholds of arousal), a number of other factors, such as music preference (Etauch & Michals, 1975), vocal versus nonvocal background music (Belsham & Harman, 1977), and musical styles (Sogin, 1988) might all interact with the role of background music in FL vocabulary learning (and learning in general). Because of its obvious pedagogical implications, a research program that systematically attempts to unravel the contribution of each of these variables is recommended.

Implications for Teaching and Conclusion

As mentioned earlier, attaching (a representation of) the newly learned FL word's form to the memory representation of the corresponding L1 word—which is essentially what successful L1-FL paired-associate learning amounts to—only concerns an initial step toward establishing a full-blown FL word representation in the FL learner's memory. As pointed out by de Groot and Van Hell (2005), two subsequent steps concern the "freeing" of the FL word form from the corresponding L1 word and "finetuning" (or "restructuring"; Jiang, 2004) the meaning associated with the newly acquired FL word such that, ideally, it is ultimately (almost) identical to the meaning native speakers of the

FL assign to this word. Similarly, Hall and Ecke (2003) assumed that, following a first step in the acquisition process where existing memory representations are exploited "parasitically" in order to establish the new FL forms in memory, more autonomous FL representations gradually develop through processes of revision and bypassing and severing earlier connections between L1 and FL representations. These autonomous representations gradually strengthen with further exposure and use (see Hall & Ecke for a detailed model of the various stages involved).

The learner can take the later steps in the acquisition process largely outside a classroom setting—for instance, by reading FL materials profusely and by communicating with native speakers of the FL to be learned. However, to reach this point of classroom-independent learning as rapidly as possible, the FL material to focus on in the classroom must be carefully chosen, structured, and sequenced. The present study has provided a set of results that might guide the choice of materials and their structuring and sequencing. The importance of introducing especially high-frequency words early on in the curriculum has already been pointed out earlier (see the introduction to this article) and, indeed, curriculum designers typically take the frequency variable into account (see Meara, 1993, and Nation, 1993, for discussions). As pointed out by Lotto and de Groot (1998), it is a pleasant coincidence that frequent words, which are especially useful to the learner because of the coverage they provide, also happen to be words that are relatively easy to learn. To promote fast initial progress further, relatively many (frequent) FL words that refer to concrete entities and have a form typical to the L1 should be introduced early on in the program. Atypical and/or abstract FL words that—because of their frequent use in the FL—are, nevertheless, indispensable early on in the curriculum require more paired-associate training than typical and/or concrete FL words. This is because the present study has shown that learning proceeds more slowly for the former FL words and that the memory representations established for them are less stable than those of typical and/or concrete FL words. De Groot and Keijzer

(2000) proposed that this might be effectuated by presenting the different types of word blockwise in the textbooks, at the same time visually marking the level of difficulty of the various groups and training the learners to pay special attention to and relearn more often the groups marked as difficult.

Finally, the present effects of background music are promising enough to warrant new research efforts that aim to map the relevant factors (e.g., the role of arousal thresholds of the brain, but also the effects of music preference, musical styles, and whether vocal or nonvocal music is presented).

Revised version accepted 14 October 2005

References

Ard, J., & Homburg, T. (1992). Verification of language transfer. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (Rev. ed.; pp. 48–70). Amsterdam: Benjamins.

Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, *96*, 124–129.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database* (CD-ROM). Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*, 586–595.

Belsham, R. L., & Harman, D. W. (1977). Effect of vocal vs. non-vocal music on visual recall. *Perceptual and Motor Skills*, *44*, 857–858.

Bogaards, P., & Laufer, B. (Eds.). (2004). *Vocabulary in a second language: Selection, acquisition, and testing*. Amsterdam: Benjamins.

Boyd Zimmerman, C. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 5–19). Cambridge, UK: Cambridge University Press.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology*, *32*, 867–873.

Daoussis, L., & McKelvie, S. J. (1986). Musical preferences and effects of music on a reading comprehension test for extraverts and introverts. *Perceptual and Motor Skills*, *62*, 283–289.

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 824–845.

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1001–1018.

de Groot, A. M. B., Dannenburg, L., & Van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, *33*, 600–629.

de Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*, 1–56.

de Groot, A. M. B., & Van Hell, J. G. (2005). The learning of foreign language vocabulary. In. J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 9–29). New York: Oxford University Press.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*, 559–617.

Etaugh, C., & Michals, D. (1975). Effects on reading comprehension of preferred music and frequency of studying to music. *Perceptual and Motor Skills*, *41*, 553–554.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: C C Thomas.

Felix, U. (1993). The contribution of background music to the enhancement of learning in suggestopedia: A critical review of the literature. *Journal of the Society for Accelerative Learning and Teaching*, *18*, 277–303.

Furnham, A., & Allass, K. (1999). The influence of musical distraction of varying complexity on the cognitive performance of extroverts and introverts. *European Journal of Personality*, *13*, 27–38.

Furnham, A., & Bradley, A. (1997). Music while you work: The differential distraction of background music on the cognitive test performance of introverts and extraverts. *Applied Cognitive Psychology*, *11*, 445–455.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.

Gathercole, S. E., & Thorn, A. S. C. (1998). Phonological short-term memory and foreign-language learning. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 141–158). Mahwah, NJ: Erlbaum.

Godden, D., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology*, *66*, 325–331.

Granger, S. (1993). Cognates: An aid or a barrier to successful L2 vocabulary development? *ITL Review of Applied Linguistics*, *99–100*, 43–56.

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL*, *40*, 69–87.

Hall, C. J., & Ecke, P. (2003). Parasitism as a default mechanism in L3 vocabulary acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *The multilingual lexicon* (pp. 71–85). Dordrecht: Kluwer Academic.

Hogben, D., & Lawson, M. J. (1997). Reexamining the relationship between verbal knowledge background and keyword training for vocabulary acquisition. *Contemporary Educational Psychology*, *22*, 378–389.

Jiang, N. (2004). Semantic transfer and development in adult L2 vocabulary acquisition. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 101–126). Amsterdam: Benjamins.

Kieras, D. (1978). Beyond pictures and words: Alternative information processing models for imagery effects in verbal memory. *Psychological Bulletin*, *85*, 532–554.

Kroll, J. F., Michael, E., & Sankaranarayanan, A. (1998). A model of bilingual representation and its implications for second language acquisition. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 365–395). Mahwah, NJ: Erlbaum.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). Basingstoke, UK: Macmillan.

Laufer, B. (1997). What's in a word that makes it hard or easy: Some intralexical factors that affect the learning of words. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary description, acquisition, and pedagogy* (pp. 140–155). Cambridge, UK: Cambridge University Press.

Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition*, *15*, 35–48.

Lawson, M. J., & Hogben, D. (1998). Learning and recall of foreign-language vocabulary: Effects of keyword strategy for immediate and delayed recall. *Learning and Instruction*, *8*, 179–194.

Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, 31–69.

McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method in context. *Journal of Educational Psychology*, *76*, 598–609.

Meara, P. (1993). The bilingual lexicon and the teaching of vocabulary. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 279–297). Amsterdam: Benjamins.

Nation, P. (1993). Vocabulary size, growth, and use. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 115–134). Amsterdam: Benjamins.

Paivio, A. (1968). A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning and Verbal Behavior*, *7*, 41–49.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. New York: Oxford University Press.

Paivio, A., & Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Canadian Journal of Psychology*, *34*, 388–399.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph*, *76*(Suppl.), 1–25.

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*, 331–347.

Rolf, P. C., & Van Rijnsoever, R. J. (1984). *Positionele letter-frekwenties van het Nederlands* [Positional letter frequencies of Dutch]. Lisse, The Netherlands: Swets & Zeitlinger.

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*, 419–440.

Schreuder, R. (1987). *Het mentale lexicon* [The mental lexicon]. Inaugural address, University of Nijmegen, The Netherlands.

Schuster, D. H. (1985). The effect of background music on learning words. *Journal of the Society for Accelerative Learning and Teaching*, *10*, 21–37.

Segalowitz, N. (1997). Individual differences in second language acquisition. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 85–112). Mahwah, NJ: Erlbaum.

Sjarbaini, L. (1998). *Het concreetheidseffect bij het leren van vocabulaire in een vreemde taal* [The concreteness effect in learning vocabulary in a second language]. Unpublished master's thesis, University of Amsterdam, The Netherlands.

Sogin, D. W. (1988). Effects of three different musical styles of background music on coding by college-age students. *Perceptual and Motor Skills*, *67*, 275–280.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*, 45–58.

Uit den Boogaart, P. C. (Eds.). (1975). *Woordfrequenties in geschreven en gesproken Nederlands* [Word frequencies in written and spoken Dutch]. Utrecht: Oosthoek, Scheltema, & Holkema.

Van den Brink, R. C. L. (2000). *De invloed van uitspreekbaarheid en klassieke achtergrondmuziek op vocabulaire acquisitie in een vreemde taal* [The influence of pronounceability and classical background music on vocabulary acquisition in a foreign language]. Unpublished master's thesis, University of Amsterdam, The Netherlands.

Van Hell, J. G., & Candia Mahn, A. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, *47*, 507–546.

Wang, A. Y., & Thomas, M. H. (1995). The effect of imagery-based mnemonics on the long-term retention of Chinese characters. In B. Harley (Ed.), *Lexical issues in language learning* (pp. 167–183). Ann Arbor, MI: Language Learning Research Club.

Zipf, G. K. (1968). *The psycho-biology of language: An introduction to dynamic philology*. Cambridge, MA: MIT Press. (Original work published in 1935 by Houghton Mifflin Company)

Appendix

Stimulus Materials

	Frequency score	Imageability score	Typicality score
Frequent/concrete/typical			
kamer (room) niek	4.19	6.00	59,704 (1,493)
mond (mouth) daak	3.98	6.52	47,594 (11,899)
vader (father) moraan	4.19	6.24	51,895 (8,649)
straat (street) gaden	3.92	6.27	44,263 (8,853)
moeder (mother) waler	4.40	6.56	36,896 (7,379)
voet (foot) balat	3.98	6.56	32,490 (6,498)
jongen (boy) srode	4.18	6.44	24,873 (4,975)
water (water) hoois	4.19	6.48	26,628 (5,326)

Appendix: (continued)

	Frequency score	Imageability score	Typicality score
Mean	4.13	6.38	40,543 (6,884)
SD	0.16	0.20	12,403 (3,109)
Frequent/abstract/typical			
manier (manner) taar	4.20	1.73	65,514 (16,379)
geval (case) helet	4.36	1.96	51,177 (10,235)
eeuw (century) mije	3.99	2.77	46,281 (11,570)
reden (reason) wons	3.98	2.16	32,561 (8,140)
kracht (power) geran	3.96	3.12	38,067 (7,613)
waarde (value) bondas	3.83	2.56	37,505 (6,251)
deel (part) baade	4.22	3.23	27,142 (5,428)
ding (thing) zaras	4.20	2.65	25,208 (5,042)
Mean	4.09	2.52	40,432 (8,832)
SD	0.18	0.54	13,434 (3,799)
Infrequent/concrete/typical			
tabak (tobacco) giedar	2.75	6.16	34,045 (6,809)
vlieg (fly) meek	2.89	6.35	56,398 (14,100)
haai (shark) kodet	2.17	6.23	46,520 (9,304)
stier (bull) voje	2.74	6.36	41,834 (10,459)
wiel (wheel) wreeg	2.95	6.27	36,800 (7,360)
pijl (arrow) hedde	2.84	6.44	31,765 (6,353)
parel (pearl) beras	2.61	6.28	30,980 (6,196)
viool (violin) gelis	2.70	6.42	27,387 (5,477)
Mean	2.71	6.31	38,216 (8,257)
SD	0.24	0.10	9,586 (2,893)
Infrequent/abstract/typical			
blaam (blame) gadet	1.89	1.19	46,429 (9,286)
lening (loan) soren	2.62	2.85	47,046 (9,409)
wrok (grudge) aalnee	2.52	1.88	57,909 (9,652)
dwang (force) lilen	2.71	2.00	39,234 (7,847)
reuk (smell) wedee	2.36	2.96	45,186 (9,037)
ruil (exchange) nans	2.81	3.08	37,367 (9,342)
wraak (revenge) horin	2.88	3.12	33,498 (6,700)
biecht (confession) woofs	2.34	2.69	27,431 (5,486)
Mean	2.52	2.47	41,763 (8,345)
SD	0.32	0.70	9,423 (1,527)

Appendix (continued)

	Frequency score	Imageability score	Typicality score
Frequent/concrete/atypical			
vrouw (woman) isfo	4.58	6.50	1,275 (319)
hoofd (head) pjubo	4.36	6.24	2,398 (480)
boek (book) ajwuc	4.21	6.54	3,552 (710)
geld (money) isiba	4.07	6.56	3,626 (725)
meisje (girl) ufhio	4.18	6.32	3,845 (769)
stad (town) ugfav	4.14	6.40	4,455 (891)
hand (hand) uzpipa	4.64	6.62	5,345 (891)
auto (car) nupwiz	3.94	6.62	9,806 (1634)
Mean	4.27	6.48	4,288 (802)
SD	0.24	0.14	2,547 (389)
Frequent/abstract/atypical			
kans (chance) ibwa	3.93	1.88	1,240 (310)
gevoel (feeling) iffux	4.03	2.72	2,146 (429)
indruk (impression) fsiwi	3.82	2.12	3,142 (628)
tijd (time) fufuz	4.66	3.08	3,606 (721)
macht (power) njopvi	3.91	2.50	3,996 (666)
einde (end) akhib	3.82	3.12	5,282 (1056)
vraag (question) azhab	4.30	2.58	6,510 (1302)
vorm (form) agowi	4.15	3.50	7,303 (1461)
Mean	4.08	2.69	4,153 (822)
SD	0.29	0.54	2,091 (411)
Infrequent/concrete/atypical			
erwt (pea) umvo	2.30	6.31	605 (151)
bruid (bride) izubi	2.66	6.20	1,654 (331)
peper (pepper) psiz	2.83	6.36	3,476 (869)
piraat (pirate) asfoi	2.20	6.20	3,601 (720)
naald (needle) ohvu	2.82	6.38	3,737 (934)
kalf (calf) obfa	2.49	6.27	3,981 (995)
kaak (jaw) cixoz	2.91	6.44	6,242 (1,248)
zeep (soap) okivoi	2.83	6.31	7,038 (1,173)
Mean	2.63	6.31	3,792 (803)
SD	0.27	0.08	2,119 (387)
Infrequent/abstract/atypical			
rente (rent) uspo	2.62	2.73	564 (141)
deugd (virtue) ubbuh	2.84	1.31	1,919 (384)

Appendix: (continued)

	Frequency score	Imageability score	Typicality score
spraak (speech) isboj	2.38	3.20	2,238 (448)
spijt (regret) puwz	2.68	1.92	3,502 (876)
mijl (mile) pupa	2.66	3.04	3,894 (974)
vloek (curse) ugowov	2.60	3.12	4,540 (757)
zwakte (weakness) pifbu	2.46	2.36	5,538 (1,108)
heiden (heathen) ibivob	2.58	2.12	7,001 (1,167)
Mean	2.60	2.48	3,650 (732)
SD	0.14	0.67	2,076 (371)

Note. Frequency scores concern the logarithm of the L1 (Dutch) word's frequency in the CELEX Dutch corpus (42.5 million words). Imageability scores concern the average score of 26 participants who rated Dutch words on imageability on a 7-point scale (1 = the word does not arouse a mental image or only with great difficulty; 7 = the word evokes a mental image quickly and easily). Words high on imageability are usually concrete; words low on imageability are usually abstract. The typicality scores concern the cumulative positional letter frequency of the nonwords in the translation pairs. The English translations of the Dutch words in the translation pairs and the average positional letter frequency per letter of the nonwords are provided in parentheses.